

1 TGCTGGGGCA CCTGAAGGAG ACTTGGGGC ACCCGCGTCG TGCCTCTGG
 51 GTTGTGAGGA GTCGCCGCTG CCGCCACTGC CTGTGCTTCA TGAGGAAGAT
 101 GCTCGCCGCC GTCTCCCGCG TGCTGTCTGG CGCTTCTCAG AAGCCGGCAA
 151 GCAGAGTGCT GGTAGCATCC CGTAATTGG CAAATGATGC TACATTTGAA
 201 ATTAAGAAAT GTGACCTTCA CCGGCTGGAA GAAGGCCCTC CTGTCACAAC
 251 AGTGCTCACC AGGGAGGATG GGCTCAAATA CTACAGGATG ATGCAGACTG
 301 TAGGCCGAAT GGAGTTGAAA GCAGATCAGC TGATGATAAACAA GAAAATTATT
 351 CGTGGTTCT GTCACTTGT TGATGGTCAG TTTCTCCCTC CTCTAACACA
 401 GGAAGCTTGC TGTGTGGGGC TGGAGGCCGG CATCAACCCC ACAGACCATC
 451 TCATCACAGC CTACGGGGCT CACGGCTTTA CTTCACCCG GGGCCCTTCC
 501 GTCCGAGAAA TTCTCGCAGA GCTTACAGGA CGAAAAGGAG GTTGTGCTAA
 551 AGCGAAAGGA GGATCGATGC ACATGTATGC CAAGAACCTTC TACGGGGCA
 601 ATGGCCTCGT GGGAGCGCAG GTGCCCCCTGG GCGCTGGGAT TGCTCTAGCC
 651 TGTAAGTATA ATGGAAAAGA TGAGGTCTGC CTGACTTTAT ATGGCGATGG
 701 TGCTGCTAAC CAGGGCCAGA TATTGCAAGC TTACAACATG GCAGCTTTGT
 751 GGAAATTACC TTGTATTTC ATCTGTGAGA ATAATCGCTA TGGAAATGGGA
 801 ACGTCTGTTG AGAGAGCGGC AGCCAGCACT GATTACTACA AGAGAGGGGA
 851 TTTCATTCTC GGGCTGAGAG TGGATGGAAT GGATATCCTG TCGCTCCGAG
 901 AGGCAACAAG GTTTGCTGCT GCCTATTGTA GATCTGGGAA GGGGCCATC
 951 CTGATGGAGC TGCACTTA CCGTTACAC GGACACAGTA TGAGTGACCC
 1001 TGGAGTCAGT TACCGTACAC GAGAAGAAAT TCAGGAAGTA AGAAGTAAGA
 1051 GTGACCCCTAT TATGCTTCTC AAGGACAGGA TGGTGAACAG CAATCTGCC
 1101 AGTGTGGAAG AACTAAAGGA AATTGATGTG GAAGTGAGGA AGGAGATTGA
 1151 GGATGCTGCC CAGTTGCCA CGGCCGATCC TGAGCCACCT TTGGAAGAGC
 1201 TGGGCTACCA CATCTACTCC AGCGACCCAC CTTTGAAAGT TCGTGGTGCC
 1251 AATCAGTGGT TCAAGTTAA GTCAGTCAGT TAAGGGGAGG AGAAGGAGAG
 1301 GTTATACCTT CAGGGGGCTA CCAGACAGTG TTCTCAACTT GGTTAAGGAG
 1351 GAAGAAAACC CAGTCAATGA AATTCAATGA AATTCTTGGG AACTTCCATT
 1401 AAGTGTGTTAG ATTGAGCAGG TAGTAATTGC ATGCAGTTTG TACATTAGTG
 1451 CATTAAAAGA TGAATTATTG AGTGCTTAAA AAAAAAAA AAAAAAAA
 1501 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAA AAAAAA (SEQ ID NO:1)

FEATURES:

5'UTR: 1-89
 Start Codon: 90
 Stop Codon: 1281
 3'UTR: 1284

Homologous proteins:

Top 10 BLAST Hits

		Score	E
CRA 18000004925454	/altid=gi 387011 /def=gb AAA60055.1 (J03503...	846	0.0
CRA 18000004920128	/altid=gi 4505685 /def=ref NP_000275.1 pyru...	793	0.0
CRA 18000004938217	/altid=gi 6679261 /def=ref NP_032836.1 pyru...	783	0.0
CRA 18000004939896	/altid=gi 66035 /def=pir DERTP1 pyruvate de...	782	0.0
CRA 18000004949905	/altid=gi 129064 /def=sp P26284 ODPA_RAT PYR...	779	0.0
CRA 18000004885327	/altid=gi 266686 /def=sp P29804 ODPA_PIG PYR...	777	0.0
CRA 18000004969398	/altid=gi 448580 /def=prf 1917268A pyruvate...	729	0.0
CRA 18000005012775	/altid=gi 1079460 /def=pir A49360 pyruvate ...	718	0.0
CRA 18000004884262	/altid=gi 1709452 /def=sp P52900 ODPA_SMIMA ...	709	0.0
CRA 18000004925713	/altid=gi 4885543 /def=ref NP_005381.1 pyru...	680	0.0

FIGURE 1A

BLAST hits to dbEST :

```
gi|10991237 /dataset=dbest /taxon=96... 1354 0.0
gi|14051054 /dataset=dbest /taxon=960... 1415 0.0
gi|14076211 /dataset=dbest /taxon=960... 1382 0.0
gi|11251518 /dataset=dbest /taxon=96... 1340 0.0
gi|13914836 /dataset=dbest /taxon=960... 1298 0.0
gi|2539160 /dataset=dbest /taxon=9606 ...1037 0.0
gi|3214685 /dataset=dbest /taxon=9606 ...1015 0.0
gi|5933458 /dataset=dbest /taxon=9606 ... 955 0.0
gi|4988948 /dataset=dbest /taxon=9606 ... 842 0.0
gi|4900594 /dataset=dbest /taxon=9606 ... 856 0.0
gi|4534604 /dataset=dbest /taxon=9606 ... 819 0.0
gi|7455087 /dataset=dbest /taxon=9606... 789 0.0
```

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

Expression information from BLAST dbEST hits:

```
gi|10991237 Neuronal precursor cells-teratocarcinoma
gi|14051054 skin
gi|14076211 skin melanotic melanoma, high MDR (cell line)
gi|11251518 muscle rhabdomyosarcoma
gi|13914836 brain neuroblastoma, cell line
gi|2539160 whole brain
gi|3214685 breast
gi|5933458 stomach
gi|4988948 pancreas - adenocarcinoma
gi|4900594 uterus - serous papillary carcinoma, high grade
gi|4534604 brain - anaplastic oligodendrogloma
gi|7455087 colon - moderately-differentiated adenocarcinoma
```

Tissue source of cDNA clone:

Fetal whole brain

FIGURE 1B

1 MRKMLAAVSR VLSGASQKPA SRVLVASRNF ANDATFEIKK CDLHRLEEGP
 51 PVTVLTRD GLKYYRMMQT VRRMELKADQ LYKQKIIRGF CHLCDGQFLL
 101 PLTQEACCVG LEAGINPTDH LITAYRAHGF TFTRGLSVRE ILAELTGRKG
 151 GCAKAKGGSM HMYAKNFYGG NGIVGAQVPL GAGIALACKY NGKDEVCLTL
 201 YGDGAANQQQ IFEAYNMAAL WKLPCIFICE NNRYGMGTSV ERAAASTDYY
 251 KRGDFIPGLR VDGMDILCVR EATRFAAAYC RSGKGPILME LQTYRYHGHS
 301 MSDPGVSYRT REEIQEVRSK SDPIMLLKDR MVNSNLASVE ELKEIDVEVR
 351 KEIEDAAQFA TADPEPPLLEE LGYHIYSSDP PFEVRGANQW IKFKSVS (SEQ ID NO:2)

FEATURES:

Functional domains and key regions:

[1] PDOC00005 PS00005 PKC_PHOSPHO_SITE
 Protein kinase C phosphorylation site

Number of matches: 7

1	16-18	SQK
2	70-72	TVR
3	137-139	SVR
4	146-148	TGR
5	282-284	SGK
6	293-295	TYR
7	307-309	SYR

[2] PDOC00006 PS00006 CK2_PHOSPHO_SITE
 Casein kinase II phosphorylation site

Number of matches: 7

1	57-60	TRED
2	137-140	SVRE
3	238-241	TSVE
4	300-303	SMSD
5	310-313	TREE
6	319-322	SKSD
7	338-341	SVEE

[3] PDOC00008 PS00008 MYRISTYL
 N-myristoylation site

Number of matches: 7

1	110-115	GLEAGI
2	114-119	GINPTD
3	151-156	GCAKAK
4	172-177	GIVGAQ
5	181-186	GAGIAL
6	183-188	GIALAC
7	235-240	GMGTSV

[4] PDOC00009 PS00009 AMIDATION
 Amidation site

146-149 TGRK

[5] PDOC00016 PS00016 RGD
 Cell attachment sequence

252-254 RGD

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	169	189	1.097	Certain

FIGURE 2A

BLAST Alignment to Top Hit:

```
>CRA|18000004925454 /altid=gi|387011 /def=gb|AAA60055.1| (J03503)
    pyruvate dehydrogenase E1-alpha precursor [Homo sapiens]
    /org=Homo sapiens /taxon=9606 /dataset=nraa /length=414
    Length = 414

Score = 846 bits (2163), Expect = 0.0
Identities = 411/421 (97%), Positives = 411/421 (97%)
Frame = +3

Query: 18 ETWGHPRRASVVRSRRCRHCLCFMRKMLAAVSRLSGASQKPASRVLASRNFANDATF 197
        ETWGHPRRASVVRSRRCRHCLCFMRKMLAAVSRLSGASQKPASRVLASRNFANDATF
Sbjct: 1 ETWGHPRRASVVRSRRCRHCLCFMRKMLAAVSRLSGASQKPASRVLASRNFANDATF 60

Query: 198 EIKKCDLHRLEEGPPVTTVLTREDGLKYYRMMQTVRRLMELKADQLYKQKIIIRGFCHLCDG 377
        EIKKCDLHRLEEGPPVTTVLTREDGLKYYRMMQTVRRLMELKADQLYKQKIIIRGFCHLCDG
Sbjct: 61 EIKKCDLHRLEEGPPVTTVLTREDGLKYYRMMQTVRRLMELKADQLYKQKIIIRGFCHLCDG 120

Query: 378 QFLLPLTQEACCVGLEAGINPTDHLITAYRAHGFTRGLSVREILAELETGRKGCAKAK 557
        Q      EACCVGLEAGINPTDHLITAYRAHGFTRGLSVREILAELETGRKGCAK K
Sbjct: 121 Q-----EACCVGLEAGINPTDHLITAYRAHGFTRGLSVREILAELETGRKGCAKAK 173

Query: 558 GGSMHHMYAKNFYGGNGIVGAQVPLGAGIALACKYNGKDEVCLTLYGDGAAANQQIFEAYN 737
        GGSMHHMYAKNFYGGNGIVGAQVPLGAGIALACKYNGKDEVCLTLYGDGAAANQQIFEAYN
Sbjct: 174 GGSMHHMYAKNFYGGNGIVGAQVPLGAGIALACKYNGKDEVCLTLYGDGAAANQQIFEAYN 233

Query: 738 MAALWKLPCIFICENNRYGMGTTSVERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFA 917
        MAALWKLPCIFICENNRYGMGTTSVERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFA
Sbjct: 234 MAALWKLPCIFICENNRYGMGTTSVERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFA 293

Query: 918 AAYCRSGKGPILMELQTYRYHGHSMSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNL 1097
        AAYCRSGKGPILMELQTYRYHGHSMSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNL
Sbjct: 294 AAYCRSGKGPILMELQTYRYHGHSMSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNL 353

Query: 1098 ASVEELKEIDVEVRKEIEDAAQFATADPEPPLLEELGYHIYSSDPPFEVRGANQWIKFKSV 1277
        ASVEELKEIDVEVRKEIED AQFA ADPEPPLLEELGYHIYSSDPPFEVRGANQWIKFKSV
Sbjct: 354 ASVEELKEIDVEVRKEIEDPAQFAAADPEPPLLEELGYHIYSSDPPFEVRGANQWIKFKSV 413

Query: 1278 S 1280
        S
Sbjct: 414 S 414 (SEQ ID NO:4)

>CRA|18000004920128 /altid=gi|4505685 /def=ref|NP_000275.1| pyruvate
    dehydrogenase (lipoamide) alpha 1; Pyruvate
    dehydrogenase, E1-alpha polypeptide-1 [Homo sapiens]
    /org=Homo sapiens /taxon=9606 /dataset=nraa /length=390
    Length = 390

Score = 793 bits (2025), Expect = 0.0
Identities = 389/397 (97%), Positives = 389/397 (97%)
Frame = +3

Query: 90 MRKMLAAVSRLSGASQKPASRVLASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED 269
        MRKMLAAVSRLSGASQKPASRVLASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED
Sbjct: 1 MRKMLAAVSRLSGASQKPASRVLASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED 60

Query: 270 GLKYYRMMQTVRRLMELKADQLYKQKIIIRGFCHLCDGQFLLPLTQEACCVGLEAGINPTDH 449
        GLKYYRMMQTVRRLMELKADQLYKQKIIIRGFCHLCDGQ      EACCVGLEAGINPTDH
Sbjct: 61 GLKYYRMMQTVRRLMELKADQLYKQKIIIRGFCHLCDGQ-----EACCVGLEAGINPTDH 113
```

FIGURE 2B

Query: 450 LITAYRAHGFTRGLSVREILAELTGRGGCAKAKGGSMHMYAKNFYGGNGIVGAQVPL 629
 LITAYRAHGFTRGLSVREILAELTGRGGCAK KGGSMHMYAKNFYGGNGIVGAQVPL
 Sbjct: 114 LITAYRAHGFTRGLSVREILAELTGRGGCAKAKGKGGSMHMYAKNFYGGNGIVGAQVPL 173

 Query: 630 GAGIALACKYNGKDEVCLTLYGDGAANQQQIFEAYNMAALWKLPCIFCENNRYGMGTSV 809
 GAGIALACKYNGKDEVCLTLYGDGAANQQQIFEAYNMAALWKLPCIFCENNRYGMGTSV
 Sbjct: 174 GAGIALACKYNGKDEVCLTLYGDGAANQQQIFEAYNMAALWKLPCIFCENNRYGMGTSV 233

 Query: 810 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS 989
 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS
 Sbjct: 234 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS 293

 Query: 990 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 1169
 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA
 Sbjct: 294 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 353

 Query: 1170 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFKSVS 1280
 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFKSVS
 Sbjct: 354 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFKSVS 390 (SEQ ID NO:5)

>CRA|18000004938217 /altid=gi|6679261 /def=ref|NP_032836.1| pyruvate
 dehydrogenase E1alpha subunit [Mus musculus] /org=Mus
 musculus /taxon=10090 /dataset=nraa /length=390
 Length = 390

Score = 783 bits (1999), Expect = 0.0
 Identities = 382/397 (96%), Positives = 387/397 (97%)
 Frame = +3

Query: 90 MRKMLAAVSRLVSGASQKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED 269
 MRKMLAAVSRLV+G++QKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED
 Sbjct: 1 MRKMLAAVSRLVLAGSAQKPASRVLVASRNFANDATFEIKKCDLHRLEEGPPVTTVLTRED 60

 Query: 270 GLKYRMMQTVRMELKADQLYKQKIIIRGFCHLCQFLLPLTQEACCVGLEAGINPTDH 449
 GLKYRMMQTVRMELKADQLYKQKIIIRGFCHLCQ-----EACCVGLEAGINPTDH
 Sbjct: 61 GLKYRMMQTVRMELKADQLYKQKIIIRGFCHLCQ-----EACCVGLEAGINPTDH 113

 Query: 450 LITAYRAHGFTRGLSVREILAELTGRGGCAKAKGGSMHMYAKNFYGGNGIVGAQVPL 629
 LITAYRAHGFTRGL VR ILAELTGR+GGCAK KGGSMHMYAKNFYGGNGIVGAQVPL
 Sbjct: 114 LITAYRAHGFTRGLPVRAILAEATGRGGCAKGKGGSMHMYAKNFYGGNGIVGAQVPL 173

 Query: 630 GAGIALACKYNGKDEVCLTLYGDGAANQQQIFEAYNMAALWKLPCIFCENNRYGMGTSV 809
 GAGIALACKYNGKDEVCLTLYGDGAANQQQIFEAYNMAALWKLPCIFCENNRYGMGTSV
 Sbjct: 174 GAGIALACKYNGKDEVCLTLYGDGAANQQQIFEAYNMAALWKLPCIFCENNRYGMGTSV 233

 Query: 810 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATRFAAAYCRSGKGPILMELQTYRYHGHS 989
 ERAAASTDYYKRGDFIPGLRVDGMDILCVREAT+FAAACRSGKGPILMELQTYRYHGHS
 Sbjct: 234 ERAAASTDYYKRGDFIPGLRVDGMDILCVREATKFAAAYCRSGKGPILMELQTYRYHGHS 293

 Query: 990 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 1169
 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA
 Sbjct: 294 MSDPGVSYRTREEIQEVRSKSDPIMLLKDRMVNSNLASVEELKEIDVEVRKEIEDAAQFA 353

 Query: 1170 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFKSVS 1280
 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFKSVS
 Sbjct: 354 TADPEPPLEELGYHIYSSDPPFEVRGANQWIKFKSVS 390 (SEQ ID NO:6)

FIGURE 2C

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF00676	Dehydrogenase E1 component	598.5	4e-176	1
PF01579	Domain of unknown function	3.0	2.3	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
PF01579	1/1	28	46 ..	153	173 .]	3.0	2.3
PF00676	1/1	66	369 ..	1	327 []	598.5	4e-176

FIGURE 2D

1 AGTTGTCCT TCTAACCCAT TGATTGTTC AATCATGTAT TTAAGTAGGA
 51 CCTATATTT ACTTGTTCCT TGCTATATCT TCAGTGTGTA GTACAGTGT
 101 TGACACAAAA TCGGTGCTCA ATAATAGGTG TTGGATGAAT GAGCAAATGA
 151 ATGAATGAAT TCATATTCTA ATGGCCTACA GAGTTCCCCT ACATGCACAA
 201 CCAATATCAC CACCCCGTGG AGATGACTCC CAAATTAAATA TTTTTAGCAA
 251 ATGTTCCAGA CTTACAACTC CAACTTCCCG GGGGACATCT TCAGATAGCT
 301 GTGCCACTGC CACCAACCAGG TCAACATGTC CCAAACCAATT CAGACCAGCT
 351 TTTTCTCTG AGCTGGACAT CTGGCCTACA ACCTTTCTAT TCTCTTTTAC
 401 CTTCTCATATT CTATCAGCAG CAGCAGCTGC TGAATATCATA CCATGCAAGT
 451 TTCTCACGTC CATCTCTGCC TTTTAATGGC GCCCTCTCAC TCCCTTAAGA
 501 AGTTTCTTC CACTGCAACCA CGATCTCTCA GTCCAGAGTC TGGCCCAGTG
 551 CCCAAATTAT TTCTCTAGCT ATGCTGAGAG CTGGTCATGC TTTGAACCTC
 601 TGCTTGAAT ACTTTCACTG ACACTGGGAG AGAATTATCT CATTGGACCA
 651 TTGTCATTGT TAGAAAATTC ATTGTATGTC TGAATGAAA TGATTTTATT
 701 CACACACACA CACACACACA CACAAAATAG CTCTTCCCTCC TGGAACATGA
 751 CTGGCCTGAA AATGTGTGAA GACATATCCA ATCCTCTCTG GTTTTACTGT
 801 TCATCCAATT TTCTGTTCTC CTCCCTGGCAG GAGGATTATA TTTCACCTTG
 851 TGGAACTCAG ACATGGTCGG GTAACCTAGCT CTGGTCCGTG AAAATTGAGA
 901 GGAAGTGACA TGTGTCACCT CTGGGAGAA GCTTGAGAG CCGGTTAAA
 951 TGATCCCTT TCTCTTCATC CATGAGACAA GCTAAGTTCC AGAGAGAGGG
 1001 TGCCACGCTG TGAGGGACCT GTGTTACGAG TACGATGGCT CGCGTCACCT
 1051 CAAATTCTTG AAATCACTGA AATTGGAGG TCAGTTGTTA CATCATAACC
 1101 CAGCCAATTC TAGTTAGCCT GTTTCTTCTC TAACCTCTTT AATCGTTCTT
 1151 CATAAGTCAC AATCGCAGCC CCTCACCGTT CTGACCACTG TCCCCCTGGAT
 1201 TCCACTCAGT TTACTCATTA TCCCCCTTA AATGTGGAGC CCAAATCTGA
 1251 ACCCGAACCC CCAGGTGCAA TCCCCTAGG ACACAAACACA ATGGGTTCTC
 1301 GAGCCCTTGT ATCCCTGAA TAGAGCCCT GTGTTGTTTG GTGTTTTGTC
 1351 TCTGTGTGTG CTTTATCAT CGGCTGAGCC ACGCTGTTAA CTCGCAGTGA
 1401 GCCTGTGAAC CAATAACTAG AGAAAAAAAGA TTTTTCCCAT TGTCCCTCTG
 1451 ACATATATTG GGAAACAAAT TTTTGATCC GGTTCAAGT AGACAGGGCA
 1501 GAACTGTCCA ACTGCTACGT GATTTTAA AGACAAAGTT AGTGGCAGAC
 1551 CATTACAGA AACCAAGATGT TCTGTTTTT GGCTCTGAGC ATGCTGCTAA
 1601 TCTTCATCAT CTAGTGTACT GAACGAGATG TACTGAACGA GGGCTGCAGA
 1651 GCTGCAGCAC CGGCAGGAGT AGGCCTCGG TAGGACGGGG CCTGCACAAAC
 1701 CTCCCCGGTA GTCAGCAGAG CGGAATCTAG GAAGGCTCCT TTCCCGGGC
 1751 GCCCTGGAGG CGGGGGCCCC ACCTTCCAC GCAGCGCCTA TCAAGCCCCG
 1801 CCTCCTCACC CGCCCGCGGC GTGGCGTCGG AAAGAGCCCT CAGCCCCCTCC
 1851 CTCTCTGGCG CTGATACCCA ATGGGAGCC TCAGGCCTTT AGCGGGGGCG
 1901 GGGCACCCCC TGGACGCGGT TCTGGTTGGC CGCGGGGGCG GCGCAGGGCA
 1951 TGACGTTATT ACGACTCTGT CACGCCCGG TGCGACTGAG GCGTGGCGTC
 2001 TGCTGGGCA CCTGAAGGAG ACTTGGGGAC ACCCGCGTCG TGCCCTCTGG
 2051 GTTGTGAGGA GTCGCCGCTG CGGCCACTGC CTGTGCTTCA TGAGGAAGAT
 2101 GCTCGCCGCC GTCTCCCGG TGCTGCTGG CGCTTCTCAG AAGCCGTTGA
 2151 GACTCCCCGG GCGGGCCGGG ATCGGGCGCG AGTGGGGCTG AGGGGGGCC
 2201 GGAGGGCAGG GCGGGCCAGG CGGGGCCACC CAGAGGGGG TGGAAGGGCG
 2251 CAGGGGAGCC GGGGAGCCTT TACTTCGCCT CGCGCCCTG CATTCCGTT
 2301 CTGGCCTCGG GAGAAGGGC ACGGACCGGG ATCACGCCAA GGTCCGTGT
 2351 AACTCCCCC TTCTGACAC CCACCTCCCG CCCCCGGGCC CAGCTGTGCG
 2401 CCAGGCGAAG TCGGTGTGCT CAAGAGGTGC CTGTTGGGTT ACAGGACACG
 2451 GAAAGGGTGG CCTCGGCCTC CTTCGAGTCT CCAATTGACC CCACTCATTT
 2501 CGGATCTTCT AACTTAATT CTCTTGACCG AGAGGCTTTG TAATAGCGTA
 2551 GAATCTGGAG ACAGGGTGGC TTCGTTCAAA CAGCACCCCTC ACCATTGACT
 2601 AGCCCTGTGA CCTTGAGCAA GTTTTAAAC GTCCCGGGGA CCCGGTTCC
 2651 TAAAATGTT GCTCGAAGTG GAGTTAATCT CTAATGGAG ATAAGAGTTA
 2701 TCTCTGAAAT GTTATCGTT ATTAAAATGT TATCAGTTAA CTCTAAAATG
 2751 GAGATAATAA GAGTCCCCAC CTCTGGGGT TGTCTTGAGG ATTCAACGAG
 2801 TGACACGTG TGAAACGATT CCAAATAGCA CCTGGCACAT AATCGATAAC
 2851 ATGTTGTTG AATAGTGTAA TTTATTGAGT CTCCAGTTCG GTATACATTT
 2901 CTTGAACACC TGTGCTCAGT TCTGAGGCGG GTTCACAGAA GGTCAAGCCTC
 2951 TTCAGAAACA AACTTCCCTCC TCTTCCCTCT CCCTCAACAT CTGAGCTTTT
 3001 CTTGGCAGTG AGTTCAAGGAG CGCCGAAGCA GAACTCAGAG GACGCTGCC
 3051 TCCCCCTCCCC TTACCTACAC ATTCTTAGGG TACAAGTAGC TAAAGCAAAG
 3101 AGCAACGATG CTTGAGGGGT GGGGGTAGA GTTGTGACT ATTTCATGGC

FIGURE 3A

3151 CTCAGCATT AGAGGTGCCT AACACCTGAG CTAGCATTCT GACCCCCCTA
3201 GGCACAGTGA GGCGTGTAA ATTGGTGTAA CTGCAGGCC CGGGATTCTG
3251 GTATTTCCCC CAGGACTTGA TACCGCTCTA CTTAGTACAG GCAAGAGATT
3301 GTCAAAAGGT AAAGAGGTAT GCCCCCTCTAG GAATCCTGTT GCCTAAAATA
3351 ATGACAAAAC TGCCGGGTGC GGTGCTCAGG CCTGTAATCC CAGCATTG
3401 GGAGGCTGAG GCAGGGAT CACCTGAAGG TCAGAAGTTC GAGATCAGCC
3451 TGGCCAACAT GGTGAAACCC CGTCTCTACT AAAAATACAA AATTAGCCGG
3501 TCGTGGTGGC GGGCTCCTGT AATCCCAGCT ACTCGGGAGG CTGAGGGGG
3551 AGAATAGCCT GAACCCGGGA GCGGAGTTG CAGTGAGCGG AGATCGGCC
3601 ATTGCACTAC GGCGCTGGCG ACAAGAAGCA AGAACACTCCG ATTTTAAAAA
3651 AAAAAAAA AAAAAAAA AAAAGCGTTC CCTTTAGGG TATCTGTGGG
3701 TAGAGGGCTG TACCGGTAGT TACGGGCTCA GAAACATCCT TCCCTTAGGC
3751 ACCTGATGTA GGTTTCTTC TTCTTCTGCA AGTCAGGTTC ATTGTTCTC
3801 GTATCAGTT GCAGGGTCCC CCCCCCCCCG CCACCTTACA GTAGGAAGAA
3851 AATTGAGTTC CAGATATGAA GTCACCTTG AAAGTGCCTA GGTATCTTC
3901 CACTTGGTGG TGTAACACT TCAGATAATT AGAAGTTTC TGTGTCACTC
3951 AACTTGTCAT GGACTAATT AGGAAACATT CCTGAAGCTT TTAAGGATAG
4001 AACTAAAAGT TTCACTTTA TTTTTTAAA GGGTGAATA ATAAACTAAC
4051 GTGTTGACTC TTTGTATTT GTAATTCTTC ATACTTATGG ATGTCTTTT
4101 ACTTAACAT AAGTAACAAA ATAGATCAAC GTTTAGTTT TTTTATATTA
4151 TACATGTAAGG AAGACATTTC GCATATAAGC CTTTCACAAA AATCTGACA
4201 GTAAACAAATA AGCAGTGGCT CACCCAAATT AGGCAGACTT ACTGCACTAG
4251 ACTCCTACCA TCTGTGTGAT ACTCCATGAA GGGAGGGAGA AGGGGAGGG
4301 GAAGGGTAGG CAGCTGGTCT GATGGCTGTG ACACAAGATA ATCCCCTTAA
4351 CCTCCCAAGA CGCTGTGTGT TTTTCTCTT TTATTCTCC CTGGTTACT
4401 TTCGTTTGTG TTGAGACAGG GTCTCTGTG CACCCAGGCT GGAGTGCAGT
4451 AGCAGGACAG CTCACTGCAG CCTTAGCCTG CTGGGCTCAA GCGATCCTCC
4501 TGCTTAGCC TCCTAGTAG CTGGGAACAC AGGCATGTGC CACCACAC
4551 CCCAGCCAAT TAAAAAAATT TTTTTTTTAC TAGAGACATG GTCTGCTAC
4601 GTGCCCAGT CTGGTCTCCA TCTCCAGGCT CAAGCAGTCC TCCCACCTCG
4651 GCCTCCAAA GTGCTGGGAT TACTCTACT CTCTTAAAC CAGGCAGGTA
4701 GGGAGATTAA TCTCAGGCTT AAAGATTGCC ATTGTCTCAT CAAAGAGTGT
4751 TTGGTGTGAA ACTTTGAAAT GAATATCAAG ATTGTGTTT TATTTTGAA
4801 TAAGGTTTAT AGTTTTCATA GTTCTTATT CATGGAAGAA GATTGAATGC
4851 ATTTAAAATG TTATTTTATT GTTGCAATT CTGTATGGCT CCTTTGTGA
4901 GATCTTTACT AGCAATGTT TGGCTTATA AGTGGTAGGT AAGAGTTTA
4951 ATTTACACTG TTAGAATCTG GAATTTTGA AACGTTTTTC CTCTTCACA
5001 TGAATGGTTC CTATGTATT AGGAAGTTAA AGTTTACTT TTTTTAATT
5051 AATTTTTTT TTTAGGCTGG AATGCAGTGG CACAGTCATA GCTCACTGTA
5101 GCCTCAGGTG TGTGCCACCA TACCTGACTA ATTTTTAAT ATTTATTTT
5151 GTAGAGATGA GAGTCTCATG TTGCCAGGC TGCGTTGAA CTCCCTGGCTT
5201 CAAGTGGTCC TCCCACCTG GCCTCCAAA GTGCTGGGG TTATAGGTGT
5251 GAGCCATCAT GCCCGGCTA GTTTTATT TTAAAATTT GAGTGGGTTG
5301 TTCGTTGCT CTGTCAGAGA GGAATCCCAT TTAACAGAGA ATCTTTTAT
5351 GGCTCTCCAG AGAAAATGAA TGGTAAACTT ATCTTTTCAA CAAGCTCTCA
5401 CTCAGAAATG ATACACACAC ACTTCTGATA GGACTTTTAG CTTCTTAAAC
5451 TTGTTCTTCT TCACTCATAT CAGTGGTTCT TATTTTTGAG ATACACAGTA
5501 ATGAAGCCAT GGGAGAAAGT ATCTAAGTAG CTTCTGGCA GTCCTAATCT
5551 TTGCAAGGCAG AAGATTACAG GCGCATGCCA CAGCACTGGG CCCCTTCTT
5601 CTCTTATTG TATAGCATT TAAGCTCTA TTGTTCAAC TCTAGGATTG
5651 AGAAAGAAGT TACCTTTCT CTGTTACTGT CGGCTGGCTG GTTGGACTC
5701 CTGCTTCCA AAAACTGCAG TTTCTGTAGT TGTTTTGGA AATTTATITC
5751 ACAATACAAT AAATTTCTGG CCCCCACAAAAT TATTTATTAAC CTGCCAAGAA
5801 TAACACATCT GTTGATTGCA TAAATATAAC CATTGATTG CTGTTTCAACC
5851 TTCTCTCAGC TTTACTTCTT CCCAAATTCC TAAATTTCTC TCACTTTTC
5901 TGAGATACAT TAGTGGACTG TCTCTGCCCTG TAAGTTAACT GAAACACTGA
5951 TTCCTAGTAT TTCAGTTGT TTCCTCCAGC ACTGTCATTG TCTGTGTTG
6001 TTGGCTTTGT CCAATAATGG TCTATTGAGG GGTGAAGATA TACGTAATTA
6051 GCTTCTGCC TATTGGCTTG TACACTCCAG GGTATACTTG GCAGATCAGT
6101 CTTAACTCTT CTCACCAAGA TCAGTCCAGT GCTGGATTAG GTAAGGTATG
6151 AACACATCAG ATGTGTTTT TATGGAGAAA TCATGTTGGT TTACACGTCA
6201 GTGTGTGAGA ATGTGGCAGA AGGGAGCTAA AATAGTATGA TAATACTACT
6251 GGATAAAATT TGTGGTCTAA CCTAAACCTT AGCCATTACA TAGAATACTT

FIGURE 3B

6301 TTGCTGTGAG CAGGTTTGCT CAGTTGTAAG ACTGGAAAGG AATCATTCT
6351 CACCCCCCGC CTCCAAGCTT TTTACCTCCA AACAGTGACA GCCACCCAAA
6401 CATCAAGAGA ACAGTGTTC AGAGAACATT TCTACTGGGG CTTCAGGAGG
6451 AGCCTGTCCA AGATTTAGGC TGTTCAAATT ATAAATTATA AAACAGCTGG
6501 CTCAGGCCA TTGTGTTAA GTCAGAGAGT GCTAAGTATC TTTTCTTTG
6551 TCTGTCTCC CTAAAGTATT TATCTCATAC TTCAATCAAT TTAAATATT
6601 TTTCTTACA GATCCAATT GATAGAAGAG TCAAGTTGC CTAGAGTGG
6651 GATTAATCA TAGTTTATT TGAAGTATAA TTTGGCTTG CTCAAAATGA
6701 ACAGTATCTG GTTATGACTA AGAATGGCAT AAAAGGCCA GACGCAGTGG
6751 CTCATGCCG CAATCCAGT ACTTTGGAG GCCAAGGCAG GTGGATCACC
6801 TGAGGTCAAG AGTTGGAGAC CAGCCTGGCC AACATGGTGA AACCCCATCT
6851 CTACTAAAAA TATAAAATT AGCCGGGCCG TGTTGGTGGG CACCTGTAAT
6901 CCCAGCTACT CGGGAGACTG AGACAGGAGA AATCACTTGA ACCCGGGAAAG
6951 CGGAGGTTGC AGTGAGCCG GATCGCACCA CTGCACTCCA GCCTGGGTGA
7001 TAAAGCAAA ACTCCGTC CAAACAAACA AACAAAAGAA TGGCATAAAC
7051 AGACACAGCT CACAGATGAT CTAGTCTCT TAGCCACTAA TTTCATTATA
7101 TTCTCACTAT AATTCTTG AAAACAAAGG ATGGGTTTGT TTTTGC
7151 TCTTGCCTG GCTTGCCTC AGATGCCGG TAATCCTGTT TCATTGCCA
7201 AAGCATGGAT TCATTTGGA GGCCAAGGAA GATGCAAACA CAGTGCACAG
7251 GGTGGAAAGAG AAGCCTATGA ATATGTTGGG GCTTATTAAA TTTCCATAAC
7301 TTCATTCTGA TAACTGATTA TTATACTTTC CAAAATAGCT GACAATTAAA
7351 AAGTACTGAT TTGTTTGTAT ATTGTTGTCT TTAAAGGCAA GCAGAGTGG
7401 GGTAGCATCC CGTAATTG CAAATGATGC TACATTGAA ATTAAGGTA
7451 GAGTGTTTA CTTTGTAAAT ATTGTTTCA CAGGTACACT CTGATATA
7501 GTTTACCTT TAGAATAGAA CATCTGATG TTCTGATTA GTCATCATT
7551 TCTTCAATAGT GTCAGGATC AGAAGTTAGT AGAAGCTTAT TCAAAGTT
7601 GGAATGTAAT TCAGTAAAT ATTGAAATAA GAAGAGTCTT AGTTGTTCT
7651 TTGAAGGTT TTTCAACCTA TAACTCAGTT GGCTTCTAGG GGCTTCTAGT
7701 GAAAATCATC TTAGAAAGAT TTCCCTCCCC CAAGCCCCAT CTCATTGCAC
7751 AGTGAGGTTT ATGGATTTA GGAACAGAGG CGATATGAAG CATTACTGAT
7801 GTGCTCCTT GCAGTTTTC AAGTCAATA TTATTGCAA TGGAGTTAGA
7851 TCTTAGAGTG GTCAACAGTG TTTGCAATGT AGTATGTTGA GGATAATAAC
7901 TACCTTATTG CATTTCAGAA ATGTGACCTT CACCGGCTGG AAGAAGGCC
7951 TCCTGTCACA ACAGTGCTCA CCAGGGAGGA TGGGCTAAA TACTACAGGA
8001 TGATGCAGAC TGTACGCCG ATGGAGTTGA AAGCAGATCA GCTGTATAAA
8051 CAGAAAATTA TTGTTGGTTT CTGTCACCTG TGTGATGGTC AGGTGAGTGG
8101 TAGGTTGTG GTGGAACGTG GTTATTAGG TACTGAAGTA TGGCTTGTAC
8151 TTATTGGCT TTACCCCTGCC ATATGTATCA GAAGAGTTTG AGGCTGGTAA
8201 TGTAAATTTC TTTTATTAT TTATTTTTT GAGACAGTCT CTCTCTGTG
8251 CCCAGGTTAG AGTACAGTGG TGATCTTGGC TCACTGCAGC CTCTGGTTAG
8301 AGTACAGTGT GATCTTGGCT CACTGCAGCC TCTGTCCACT GGGCTCAAGC
8351 AATCCTCCCA CCTCAGCCTC CCGAGTATGT GGGACACAG GTGCACACCA
8401 ACACACCCAG CTAATTGTTG TATTGTTGG AGATACGGGG TTTCACTATG
8451 TTGCCCCAGGCTAGT CTTCTGGCT CAAGTGGTCC GCCCACCTTG
8501 GCCTCCCAAG GTGCTAGGAT TACAGCGTG AGCCACTGTG CCTGGCTGAA
8551 GCCAGTATTG TAGAATTAAA AAGTAAATG CCAAAACCTG CTATGAAGCT
8601 TAGGCTAAAG AATTCAATCA CACATAACAT TGCCAGTTT CTGTACCTGT
8651 TCTTAGAGTT TTACTATTAA AAAACTTTCT GGCACATATGA TCGCCTGTAC
8701 TGTATATAAT TTGGAGAGAA AGGATTAGTT TTTTTTTGT TTTGTGGCT
8751 TAGGTCAAGG GTTAGAGTCA AATACCTACA AGGGCCAGCC AGGTAGAATA
8801 AATGAGTGAA GAAGGCTAGG TATACAAAAG AGAAAATGGT GACAGGGACT
8851 CATGCTGAAC TGGCACCAGC ATGCCCTACC CAGAGGAATG CCATGACTTG
8901 GTTCCAGCCA GTTGGTGCCA TGTGAAATC AGGGTAATG TTTCTGTG
8951 TCCATGTCTA AGAGAAGGCG GAAGTCTGGA TTTTCATGTG AAATTCCAG
9001 TGTTTAATG TTGACATCTG ATGTAGGCTT TTATTGTTAGG TCATCATACA
9051 GGAGAAAGGA AGGAAGTGGC ACATGTGTGG GTTGCCAGTT TATTGCTCT
9101 GGTITGGGCC TTCCACTCTG TATTGTTGGG AAAAATAGCT ACTTTCTCTG
9151 GTTATTAAATG ACAGGGTCTA CTAGCCCACA TATTCACTG TGGTCTAGGA
9201 AACGTTTTA TTTAGAAACCA TGTATCATAT TGCCTCATAG TTTCTCC
9251 CTCTAACACA GGAAGCTTGC TGTGTGGCC TGGAGGCCGG CATCAACCC
9301 ACAGACCATC TCATCACAGC CTACCCGGCT CACGGCTTTA CTTTCACCC
9351 GGGCTTCC GTCGGAGAAA TTCTCGCAGA GCTTACAGGT TTGCTGTG
9401 TTTACAGAAA GGGAAATGA GTGGATTAAG TTTTAAATA TCTGTGCATT

FIGURE 3C

9451 AAGATGCTAT TATGAGTTAA TATTGTTAA AAATTTAAG TTTCTTTTT
 9501 TAACCCCTCTC TCCTTGGTG CTCTGGTACT TCTGTTGTGC TCTTGAGTTA
 9551 ACTGACCATT TGTGAAGTTC TCTGGCCCT CAGGTAAAAG TTAAAACAG
 9601 GTTGGTGCTA TAAAATCACA GTAGGTTGG TTATCATTCA AGCATGCCAG
 9651 AAGAAGTCTA GCAGTCATAG AAAGTAAGTT CGGTTGAAGC ACTCCATGGT
 9701 ATGCAATGTA AATTCTAGAA ATCTTCTAA TATTCCCCTT TTCTTTGTCC
 9751 CCCGTGACTA TTGTTTGTG TTGGTGGTTT TTTTTTTTTT TTTTTTTTGA
 9801 GACTGTGCTC CACTCCGTG TCCAGGTGGT GTGAGTGGT GTGATCAGGG
 9851 CTCACTGCAA CCTCCACCTC CGGGGTCAA GTGATTCTCA TGCCCTCAC
 9901 TCCGTAGTAG CTGGGACTAC AGGCATGCAC CACCACACCT GGCTAATTT
 9951 TGATTTTA GTAGAGATGG GGTTTCAACA TGGTGGCCAG GCTGGTCTCC
 10001 AACTCCGTAC CTCAGGTGAT CCACCTGCCT TGCCCTCCCA AAGTGTGCTG
 10051 GGGTACAGG CGTGAACCCAC CGCACCTGGC CTGTTTGTGTT TTTTGAGAC
 10101 AGAGTCTCGC TTGTTGCCCC AGGCTGGAGT GCAGTGGCCT GCCTCACCC
 10151 CCCAAATGTC TAGGATTACA GGCCTGAGCC ACTGTGCCCG GTCCCTCTCC
 10201 TCCTCCCTTT TTTTTTTTTT TTTTGAGACA GAGTTTCACT CTTTCACCCA
 10251 GGCTGGAGTG GCTGGAGTGA AGTGGTATGA TTTTGGCTCA CTGCAGCCTC
 10301 CGCCCCCCGG GTTCAAGCAA TTCTCCTGCC TCAGCCTCCT GAGTAGCTAG
 10351 GATTATAGGT GCCAACCCAC CACACCTGGC TAATTCTGT ATTTTTAGTA
 10401 GAGACCAGGT TTCACCATGT TGGCCAGGCT GGTCTTGAAC TCTTGACCTC
 10451 AGGTGATCCA CCCTCTTCGG CCTCCAAAAA TGTAGGATT ACAGGCGTGA
 10501 GCCGCGTGC CGGGCCCTCC TTGACTCTTG AACTATGGTT GTCCCTCTAT
 10551 ATATCCAGGG GATTGGTTCT AGGACCCCTCG AGTATACAAA AATCCTCAAA
 10601 TACTCAAGTC CCAAAGTCAG CCTTCCATAT CTTCGGGTTT GCATCCTGAG
 10651 AATATTCTAT TTTCATACA TGTGTTGGCTG AAAAAAATC TGTGTATAAG
 10701 TGACCTGTG CAGTTCAAAC CCTGTCAAG GATTGAATAT ATTTAGTGT
 10751 CTAGTATAGG AGAGGTCCTA AGATTTTGT AACTGGCCAG AAAACCCAGA
 10801 AAAGTCCAGG GTATCATCTG GATGAAACAT CTGAAGGAAA CTAAGTGA
 10851 AGAGAGTAGG AAAAGCTGGA AAGGTTGAAG CACATGGAAC TAGTGA
 10901 ACAAGGAGAA ACATGTGTTT GCCTGGAGGG ACAGGTACTT AGACGACTGA
 10951 ACTGGCTCT GTGTTCTAAT GTTGGACCT CAGAGTACAT ATTTGGGTTG
 11001 CGGTTGGTT TGCTTGTAG AGTTGGTTT TCTGCACAT GTGTATGTT
 11051 TGCCATTTC AGGACGAAA GGAGGTTGTG CAAAGGGAA AGGAGGATCG
 11101 ATGCACATGT ATGCCAAGAA CTTCTACGGG GGCATGGCA TCGTGGGAGC
 11151 GCAGGTAGTC AAGGACGAGG ATTGTGTGCT GTCTTAGATT TGGCCCTGGA
 11201 CTTGTCTTG AAAAACCTT CACAGCCCCA GACAACCTTT CCTGAAGCTA
 11251 GTACAGCCAT GTGCTGCACA GTGACGCTT GGTCAATGTC GCATATATGA
 11301 TGTGGACCC ATAAGATTAT AATGGAGCTG AAAAATTCT CTCGCCTAGT
 11351 GATGTTGTAG TGGCACAACA CATTACCTT TCTACGTTA GGTACACAAA
 11401 TATTTGCCT ACAGGATTCA GTAGAGTCAC ATGCTGTGCA GGGTTGTAGC
 11451 CTAGGAGCAG TAGGCTCTAC TATACAGCCT AGGTGTGCAG TGGGCTGTAC
 11501 CATCTAGGTT CGTGCATTAC AGTATGGTGT TCACATGACA AAATCGCTA
 11551 GTGATGCAAT TCTGAGATA TATCCCTGTT GTAAAGTGAC GCGTGA
 11601 TTTGGGGCT TGGTTGCTT TAAAGACCT AGTGCCTCAT ATCCCTACCGT
 11651 TTGAGAGATG AGTAGATTG GATGGTGATT TATAATGTTT CTTTTAGGT
 11701 GTCTGCTGTT TTATAAGTAA GCAGGAACCT CTAGCAGTGG AGCCATACCT
 11751 TCCCTTCTCCT ATTTATATTT CAGTACATTA ATTGCTTTAT CTTGTCAACT
 11801 TCATTTGGG GTCTTGTTC TCATCAGTTA GTGAATGATG AAGAATTAAC
 11851 AGCACAAAAT TATATCCGA CTGTTCTTT TCCTTTCTAA TATATTAAGA
 11901 TTCTATTATG TGTGTTTTT TTTAAACCT AGGTTTTATT TTTCTTTTG
 11951 AAATGGAGTC TTGCTCAGCC GCCCAGGCTG GAGCAGTGGT GTAATCTCAG
 12001 CTCACTGCAA CCTCCACCCCG CGGGTCAAG CAATTCTCCT GCCTCAC
 12051 CCCGAGTAGC TGGGAATATA GTTACGTGCC ACCATGCCCA ACCATTTTT
 12101 GTATTTTAG TAGAGACGGG GTTCAACCAT CTGTCCAGG ATGGTCTCGA
 12151 TCTGTGGACC TCGTGTCTG CCCAAAGTGC TGGGATTACA GGCAGTGA
 12201 ACCACGCCCG GCCAGGTTT ATTTTTAAC TCTTGAATGC AGAAATGTTA
 12251 GTGCTTACTG GTAAAATAG AACATAGTAT TTATATATTA CTTTAGTGT
 12301 TTATTGAAAA TATCGGAGGT GGGATAAAACA GAGAGATAGG GTTGGAGGA
 12351 GAGTTTGTAG CAGCAGTGTAA ATTCTGTGT CAGATTCTGG CCAGGAGTGA
 12401 AAATGCAGGG CATTAAATTAG TATCTCCCT CATGGATTTC TGTGGTCTC
 12451 TTCTCGGTG TCCTTAATGT TAGGTGCCCT TGCGCGCTGG GATTGCTCTA
 12501 GCCTGTAAGT ATAATGGAAA AGATGAGGTC TGCCCTGACTT TATATGGCGA
 12551 TGGTGTGCT AACCAAGTAA TTATGCTCT TAACCTCCCA AAAACAGTCT

FIGURE 3D

12601 TATTTCAAA GTCTTAATA TTTACAGTTG AATTCTAAA GAAGTAGCAT
 12651 ATTGCTTATT AGGTGAAATA GCAAGCTTA TGCGTAGCTC AAATTTGGTT
 12701 GACTTATGGC CAGATTAGAG ATTGACCTCT TAGCGTTGTT TCACAAGAGA
 12751 CTTACGGGGG CACATTCCCTG TGAAGGAGCT CACCTTTGCT CTACATCAGT
 12801 GCTTGGCAAA GGCCCTGTGG TAAAGGACCT CCCCACAAACC TATTGAAAAA
 12851 CAATACAGAC CCATTCTCTT GGATGTCCGG GCTGGCAGTG TCAAATTCCG
 12901 ATAATAGCGT CTGAGTCCTA ACTCAGTTTC TATGCTTCTC TTGTTACCGA
 12951 GTAATCCCCA GTCTGTGGCC AGCACTCTGT GAAGCCCTGT TCTAGAGGCT
 13001 GATTCTTAGG TGCTGGTCA CTCTGGCTAT CCAGTGGGCC TGATAGATTT
 13051 CATAATTGATC TTTTTTCCAG TGTGTCTCTT ACTGCTAGCA TGGCCCCAAA
 13101 GAAACAAGTA GTAGTTGGTT TGTCACTTC CTTAGTTGCA AGAGTATGAT
 13151 GCCTGCTACT TCTCCTCCAC CACCCACCCCC GCTTCCCTC ACCACCCAAA
 13201 GCTCGGTTTT AGAAGAGGAG GCTTTCTGTG CTTTATGAAA GCTTTCTGTG
 13251 CCAGGCAGAG CAGCAGCTGT TAGAGATGAT GAAGCCTGGA GAAAGAACCC
 13301 AAATGAAACC CCTTTTCTGA ACTACTTCCA GGGCCAGATA TTCGAAGCTT
 13351 ACAACATGGC AGCTTGTGG AAATTACCTT GTATTTCAT CTGTGAGAAT
 13401 AATCGCTATG GAATGGGAAC GTCTGTTGAG AGAGCGGCAG CCAGCACTGA
 13451 TTACTACAAG AGAGGCGATT TCATTCTGG GCTGAGAGTA AGGACACCTG
 13501 TGGTGGGGCC GGGGCCAAGG CCAAGCCAA GGGTATGTAC CTTGTGCAGA
 13551 CCCTTGACGA TCTTAGAAC ATTGGAGAGT TTCATTCTCA TACAGGAGCA
 13601 GGTCATGTGA AAGTAAAATG GTTGGGGCA GTTGGATTCA TGCTTCGCC
 13651 CTCCCCTGTT TATTACCAAG TGGATGGAAT GGATATCCTG TGCGTCCGAG
 13701 AGGCAACAAG GTTGTCTGCC GCCTATNGTA GATCTGNNNN NNNNNNNNNNN
 13751 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 13801 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 13851 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 13901 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 13951 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14001 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14051 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14101 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14151 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14201 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14251 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14301 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14351 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14401 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14451 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14501 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14551 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14601 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14651 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14701 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14751 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14801 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14851 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14901 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 14951 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15001 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15051 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15101 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15151 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15201 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15251 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15301 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15351 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15401 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15451 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15501 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15551 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15601 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15651 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN
 15701 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN TTAGTGTAC

FIGURE 3E

15751 TTCAGATGAT ATAGGCATAA GATACATTGG TTTTGCTGGC TGTGCTTCTT
 15801 TAGGGGACT TAAGGGAGAA AGGCAAGGC CATGGATTTC CTGCTTGGCG
 15851 CTCTGATGTC TCAAAGTCTA ATTATCACCA CACACACCAT CTCTGCTGTC
 15901 CCCACCCATG TAGTATACAG GAGCCAAAT GGGTGGGACA AGTGACACTT
 15951 CTTTAGAACC TTACATCTAA ATCAAAGCAG CAAGAAAAA CTTGGCCCC
 16001 GTTGTGGTA ATGCCAGGGA AGCCATGTGA CTCACCAGTG TACGGTTTC
 16051 TAGAAAAGAC AGAACAGTT ATTACAGAAAT GTTACGGCTGC GTTCTGGTAT
 16101 TTTGAAAGTA TAACAACAAC TCTGCCACGC CTATAGTGAC ATAAGCATTG
 16151 GTATGCCCT TTGTTTCAGA AACACACTTC TGATTTTCAC CTCATTGGGA
 16201 CAATCCAACC CCATATCATG TTTCATCACG CGCTCCITGC TCTACTGGAA
 16251 CTGCTCTTAC TGATCGATTA CTACTTTCC CTCCCCATAG TTACCGTACA
 16301 CGAGAAGAAA TTCAGGAAGT AAGAAGTAAG AGTGACCCCTA TTATGCTTCT
 16351 CAAGGACAGG ATGGTGAACA GCAATCTTGC CAGTGTGGAA GAACTAAAGG
 16401 TACAGTCACT TGTTCATGGT GGTTTGAAGG TTGGCTTAA AAGTTGCAC
 16451 CCCCTGGTGG CCACAGAGTT TGTGTGGTT CCTCCAAGCC CAGAAAGTGA
 16501 TGTCTGGGA CATAAAATAGT TCCATAGTTC CAAAGTCCCT TGGGGTGGGG
 16551 GCTTTCCCT TAGTTTCCTC TATTCAAAAT TGTTTACTC TTCAAGATTTC
 16601 AGATTTGGT GGACTGTGAA CCACCATCAC AGTGGCAAAG CCCCCACAGT
 16651 AGTATGGTC TTTTTCCCTA AAAGTATACT GTGGATTTT AATTCAAAA
 16701 ATAGATACAC CCTAGAAATC TGTNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 16751 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 16801 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 16851 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 16901 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 16951 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17001 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17051 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17101 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17151 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17201 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17251 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17301 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17351 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17401 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17451 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17501 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17551 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17601 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17651 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17701 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17751 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17801 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17851 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17901 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 17951 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18001 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18051 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18101 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18151 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18201 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18251 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18301 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN
 18351 NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNNNN NNNNNNNNN

(SEQ ID NO:3)

FEATURES:

Start: 2090
 Exon: 2090-2146
 Intron: 2147-7386
 Exon: 7387-7446
 Intron: 7447-7918
 Exon: 7919-8092
 Intron: 8093-9240
 Exon: 9241-9388

FIGURE 3F

Intron: 9389-11062
 Exon: 11063-11154
 Intron: 11155-12473
 Exon: 12474-12566
 Intron: 12567-13331
 Exon: 13332-13487
 Intron: 13488-13669
 Exon: 13670-13727
 Intron: 13728-15920
 Exon: 15921-16007
 Intron: 16008-16290
 Exon: 16291-

CHROMOSOME MAP POSITION:

Chromosome X

ALLELIC VARIANTS (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
1785	G	T	Beyond ORF(5')			
1895	G	A	Beyond ORF(5')			
2118	G	C	Exon	10	R	P
5144	T	C	Intron			
7932	A	G	Exon	44	H	R
8015	C	T	Exon	72	R	C
8063	C	A	Exon	88	R	S
8066	G	A	Exon	89	G	S
9307	C	G	Exon	120	H	D
9349	C	T	Exon	134	R	W
9350	G	A	Exon	134	R	Q
11066	G	A	Exon	148	R	Q
11128	G	A	Exon	169	G	R
11135	A	G	Exon	171	N	S
11143	G	A	Exon	174	V	M
12486	G	C	Exon	182	A	P
12558	G	A	Exon	206	A	T
13376	T	C A	Exon	223	F	F L
13378	C	T	Exon	224	P	L
16233	G	C	Intron			
16354	G	A	Exon	330	R	K
16377	T	G	Exon	338	C	G

Context:

DNA Position	Sequence
1785	TCAAGTAGACAGGGCAGAACTGTCAAATGCTACGTGATCTTTAAAGACAAAGTTAGTG GCAGACCATTACAGAAAACCAGATGTTCTGCTTTGGCTCTGAGCATGCTGCTAATCTT CATCATCTAGTGTACTGAACGAGATGTA CTGAACGAGGGCTGCAGAGCTGCAGCACCGGC AGGAGTAGGCCTCGTAGGACGGGCCTGCACAACCTCCCCGGTAGTCAGCAGAGCGGA ATCTAGGAAGGCTCTTCCCGCGGCCCTGGAGGCGGGGGCCCCACCTTCCCACGCAG [G, T] CGCTATCAAGCCCCGCCTCTCACCCGCCCGCGCGTGGCGTGGAAAGAGCCCTCAGCC CCTCCCTCTGGCGCTGATACCAATGGGCAGCCTCAGGCC CCCCCTGGACGCCGTCTGGTTGGCCGCCGGCGCAGCGCATGACGTTATTACGAC TCTGTCACGCCGCCGTGCGACTGAGGCGTGGCGTCTGCTGGGGCACCTGAAGGAGACTTG GGGGCACCCCGCTCGCCTCTGGTTGTGAGGAGTCGGCGCTGCCCACTGCCGTG 1895 TGCTAACCTTCATCATCTAGTGTACTGAACGAGATGTA CTGAACGAGGGCTGCAGAGCTG CAGCACCCGGCAGGAGTAGGCCTCGTAGGACGGGGCTGCACAACCTCCCCGGTAGTC CGAGAGCGGAATCTAGGAAGGCTCTTCCCGCGGCCCTGGAGGCGGGGGCCCCACCT

FIGURE 3G

TCCCACGCAGGCCTATCAAGCCCCGCCCTCCTCACCCGCCGCGCGTGGCGTCGGAAAG
 AGCCCTCAGCCCCCTCCCTCTGGCGCTGATACCCAAATGGGCAAGCCTCAGGCCTTAGCG
 [G, A]
 GGGCGGGCACCCCTGGACGCCGTTCTGGTTGGCCGCCGGCGCAGCGCATGACG
 TTATTACGACTCTGTACGCCGGTGCGACTGAGGCCTGGCGTCTGCTGGGGCACCTGA
 AGGAGACTTGGGGCACCCGCGTCTGGCTCCTGGTTGTGAGGAGTCGCCGCTGCCGCC
 ACTGCCTGTCTTATGAGGAAGATGCTGCCCGTCTCCCGCTGCTGTCTGGCGCTT
 CTCAGAACGCCGTGAGACCTCCCGGCCGGGGATGGGGCGAGTGGGGCTGAGGCG
2118
 GGCGTGGCGTGGAAAGAGCCCTCAGCCCCCTCCCTCTGGCGCTGATACCCAAATGGGCA
 GCCTCAGGCCCTTAGGGGGGGGGCACCCCTGGACGCCGTTCTGGTTGGCCGCC
 CGCGCAGCGCATGACGTTATTACGACTCTGTACGCCGGTGCGACTGAGGCGTGGC
 GTCTGCTGGGCACCTGAAGGAGACTTGGGGCACCCGCGTCTGCTGGGTGTGA
 GGAGTCGCCGTGCCCACTGCCTGTCTTATGAGGAAGATGCTGCCGCCGTCTCCC
 [G, C]
 CGTGCTGTCTGGCGCTTCTCAGAAGCCGGTGAGACCTCCCGGCCGGGGATGGGGCG
 CGAGTGGGCTGAGGCCGGGCCGGAGGGCAGGCCGGGCCAGGCCGGCACCCAGACGG
 GGTGAAAGGCCAGGGAGGCCGGGAGCCTTA
5144
 TGAATGCATTTAAAATGTTATTTATTGTTGCATTTCTGTATGGCTCCTTTGTGAGAT
 CTTTACTAGCAATGTTTGGCTTATAAGTGTAGGTAAGAGTTTAAATTACACTGTTA
 GAATCTGAATTTTGAAACGTTTCCCTTTACATGAATGGTTCTATGTATTAGG
 AAGTTAAAGTTTACTTTTTAAATTAAATTTTTTAGGCTGGAATGCAGTGGCAC
 AGTCATAGCTCACTGTAGCCTCAGGTGTGCCACCACCTGACTAATTAAATT
 [T, C]
 ATTTTGAGAGATGAGAGTCTCATGTTGCCAGGCTGGCTTGAACCTGGCTCAAG
 TGGCCTCCCACCCGGCTCCCCAAGTGTGGGATTATAGGTGTGAGCCATCATGCC
 GGCCTAGTTTATTTAAATTGAGTGGGTTCTGGTCTGTCAAGAGAGGAA
 TCCCATTAAACAGAGAACTTTTATGGCTCCAGAGAAAATGAATGGTAAACTATCT
 TTCAACAAGCTCACTCAGAAATGATAACACACACTCTGATAGGACTTTAGCTTC
7932
 AAGAGTCTTAGTTCTTGAAGGTTCTTCAACCTATAACTCAGTTGGCTTAGGG
 GCTTTCAGTAAAATCATCTTAGAAAGATTTCTCCCAAGCCCCATCTCATTGCACA
 GTGAGGTTATGGATTAAAGGAACAGAGGCGATATGAAGCATTACTGATGTCTTTG
 CAGTTTTCAAGTTCAATATTATTGCAATGGAGTTAGATCTTAGAGTGGCAACAGTGT
 TTGCAATGTAGTATGTGGAGGATAATAACTACCTTATTCCATTTCAGAAATGTGACCTTC
 [A, G]
 CCGGCTGGAAGAAGGCCCTCTGTCACAACAGTGTCAACCAGGGAGGATGGCTCAAATA
 CTACAGGATGATGCAGACTGTACGCCGAATGGAGTTGAAGCAGATCAGCTGTATAAAC
 GAAAATTATTGTTCTGTCACTTGTGATGGTCAGGTGAGTGGTAGGTTGTGGT
 GGAACGTGTTATTTAGGTACTGAAGTATGGTTGTACTTATTGGCTTACCTGCCAT
 ATGTATCAGAACAGAGTTGAGGCTGTAATGTAATTTCATTATTATTGAA
8015
 AAAGATTCCTCCCCAAGCCCCATCTCATTGCACAGTGAGGTTATGGATTTAAGGAA
 CAGAGGCGATATGAAGCATTACTGATGTCTCTTGCAGTTTCAAGTTCAATATTAT
 TTGCAATGGAGTTAGATCTTAGAGTGTCAACAGTGTGTTCAATGTAGTATGTGGAGGAT
 AATAACTACCTTATTCCATTTCAGAAATGTGACCTTCACCGGCTGGAAAGAAGGCCCTCCT
 GTCACAAACAGTGTCAACCAGGGAGGATGGCTCAAATAACTACAGGATGATGCAGACTGTA
 [C, T]
 GCCGAATGGAGTTGAAAGCAGATCAGCTGTATAAACAGAAAATTTCAGTGGTTCTGTC
 ACTTGTGTGATGGTCAGGTGAGTGGTAGGTTGTGGTAGGTTGTGAACTGTGTTATTAGGTACTG
 AAGTATGGCTGTACTTATTGGCTTACCCGCCATATGTATCAGAACAGAGTTGAGGCT
 GTTAATGTAATTTCATTATTATTGAGACAGTCTCTCTGTGCCAG
 GTTAGAGTACAGTGGTATCTGGCTCACTGCAGCCTCTGGTAGAGTACAGTGTGATCT
8063
 GGATTTAGGAACAGAGGCGATATGAAGCATTACTGATGTGCTCTTGCAGTTTCAA
 GTTCAATATTATTGCAATGGAGTTAGATCTTAGAGTGTCAACAGTGTGCAATGTAG
 TATGTGGAGGATAATAACTACCTTATTCCATTTCAGAAATGTGACCTTCACCGGCTGGAA
 GAAGGCCCTCTGTCACAACAGTGTCAACCAGGGAGGATGGCTCAAATAACTACAGGATG
 ATGCAGACTGTACGCCGAATGGAGTTGAAAGCAGATCAGTGTATAAACAGAAAATTATT
 [C, A]
 GTGGTTCTGTCACTTGTGATGGTCAGGTGAGTGGTAGGTTGTGGTAGGAACTGTGTT
 ATTAGGTACTGAAGTATGGCTGTACTTATTGGCTTACCCGCCATATGTATCAGAA

FIGURE 3H

GAGTTTGAGGCTGGTAATGTAATTTCCTTTATTTATTTATTTGAGACAGTCTCTC
 TCTGTGCCCAAGGTTAGACTACAGTGGTACTTGGCTCACTGCAGCCTCTGGTAGAGT
 ACAGTGTGATCTGGCTACTGCAGCCTCTGCCACTGGCTCAAGCAATCCTCCACCT

8066
 TTTAAGGAACAGAGGCATATGAAGCATTACTGATGTGCCCTTGAGTTTGAACTGTATT
 CAATATTATTGCAATGGAGTTAGATCTTAGAGTGGTCAACAGTGGCAATGTAGTAT
 GTGGAGGATAATAACTACCTTATTCAGAAATGTGACCTTCACCGGCTGAAAGAA
 GGCCCTCTGTCAACACAGTGCACCCAGGGAGATGGCTAAACTACAGGATGATG
 CAGACTGTACGCCAATGGAGTTGAAAGCAGATCAGCTGTATAAACAGAAAATTATCGT
 [G, A]
 GTTTCTGTCACTTGTGTATGGTCAGGTGAGTGGTAGGTTGTGGGAACTGTGTTATT
 TAGGTACTGAAGTATGGCTTGTACTTATTGGCTTACCCGTGCAATGTATCAGAAGAG
 TTTGAGGCTGGTAATGTAATTTCCTTTATTATTATTGAGACAGTCTCTCT
 GTCGCCAGGTTAGAGTACAGTGGTATCTGGCTCACTGCAGCCTCTGGTAGAGTACA
 GTGTGATCTGGCTCACTGCAGCCTCTGCCACTGGCTCAAGCAATCCTCCACCTCAG

9307
 AATGTTGACATCTGATGTAGGTTTATTAGGTATCATACAGGAGAAAGGAAGGAAG
 TGGCACATGTGTGGGTGCCAGTTATTGCTCTGGTTGGGCTTCACTCTGTATT
 GGGGAAATAGCTACTTCTGGTTATTAAATGACAGGGCTACTAGCCCACATATT
 ACTGTGGTCTAGGAAACGTTTATTAGAAACATGTATCATATTGCTCATAGTTCTC
 CTTCCTCTAACACAGGAAGCTGTGTGGGCTGGAGGCCATCAACCCCCACAGAC
 [C, G]
 ATCTCATCACAGCTACCGGGCTCACGGCTTACTTACCCGGGCTTCCGTCCGAG
 AAATCTCGCAGGCTTACAGGTTGCTGTTGATTTACAGAAAGGGAAATGAGTGGATT
 AAGTTTAAATATCTGTGATTAAGATGCTATTATGAGTTAATATTGTTAAAAATT
 AAGTTTCTTTAACCCCTCTCCCTGGTGCTCTGGTACTCTGTGTGCTCTTGAG
 TTAACTGACCATTGTAAGTTCTGGCCCTCAGGTAAAAGTTAAAACAGGTTGGTG

9349
 CAGGAGAAAAGGAAGGAAGTGGCACATGTGTGGGTTGCCAGTTATTGCTCTGGTTGG
 CCTTCCACTCTGATTTGGGAAATAGCTACTTCTGGTTATTAAATGACAGGGCT
 TACTAGCCCACATATTCACTGTGGTCTAGGAAACGTTTATTAGAAACATGTATCAT
 ATTGCTCATAGTTCTCCCTCTAACACAGGAAGCTGTGTGGGCTGGAGGCC
 GCATCAACCCCCACAGACCATCTCATCACAGCCTACCGGGCTCACGGCTTACTTCACC
 [C, T]
 GGGGCCTTCCGTCGGAGAAATTCTCGCAGAGCTTACAGGTTGCTGTTGATTTACAGAA
 AGGGAAATGAGTGGATTAAGTTTAAATATCTGTGATTAAGATGCTATTATGAGTTA
 ATATTGTTAAAATTTAAGTTCTGGTACTGACCATTGTGAAGTTCTCTGGCCCTCAGGTAAA
 GTTTAAAACAGGTTGGTGCTATAAAACAGTAGGTTGGTTATCATTCAAGCATGCCA

9350
 AGGAGAAAAGGAAGGAAGTGGCACATGTGTGGGTTGCCAGTTATTGCTCTGGTTGG
 CCTTCCACTCTGATTTGGGAAATAGCTACTTCTGGTTATTAAATGACAGGGCT
 ACTAGCCCACATATTCACTGTGGTCTAGGAAACGTTTATTAGAAACATGTATCATA
 TTGCTCATAGTTCTCCCTCTAACACAGGAAGCTGTGTGGGCTGGAGGCC
 GCATCAACCCCCACAGACCATCTCATCACAGCCTACCGGGCTCACGGCTTACTTCACC
 [G, A]
 GGGCCTTCCGTCGGAGAAATTCTCGCAGAGCTTACAGGTTGCTGTTGATTTACAGAA
 GGGGAAATGAGTGGATTAAGTTTAAATATCTGTGATTAAGATGCTATTATGAGTTA
 TATTGTTAAAATTTAAGTTCTGGTACTGACCATTGTGAAGTTCTCTGGCCCTCAGGTAAA
 GTTTAAAACAGGTTGGTGCTATAAAACAGTAGGTTGGTTATCATTCAAGCATGCCAG

11066
 TCCTAAGATTTGTAACTGGCAGAAAACCCAGAAAAGTCCAGGGTATCATCTGGATGG
 AACATCTGAAGGAAACTAAGTGAATAGAGTAGGAAAAGCTGGAAAGGTTGAAGCACAT
 GGAACATAGTGAAGGACAAGGAGAAACATGTGTTGCCTGGAGGGACAGGTACTTAGACG
 ACTGAACCTGGCTCTGTGTTCAATGGTTGAGCCTCAGAGTACATATTGGGTGCC
 TGGTTTGCTTAGAGTTGGTTGTCAGATGTGATGTTCTGCCATTCCAGGAC
 [G, A]
 AAAAGGAGGTTGTGCTAAAGGGAAAGGAGGATCGATGCACATGTATGCCAAGAAC
 CGGGGCAATGGCATCGTGGGAGCGCAGGTAGTCAGGACGAGGATTGTGCTGCTT
 GATTGGCCCTGGACTTGTCTGAAAAACCTTCAGGCCCCAGACAACATTCTGAA
 GCTAGTACAGCCATGTGCTGCACAGTGACGCTTGGTCAATGTCGATATGATGTTGG
 ACCCATAAGATTATAATGGAGCTGAAAATCCTGTCGCCAGTGATGTTGAGTGGCAG

FIGURE 3I

11128 CATCTGAAGGAAACTAAGTGAAGTAGAGAGTAGGAAAAGCTGGAAAGGGTTGAAGCACATGG
 AACTAGTGAAGGACAAGGAGAACATGTGTTGCCTGGAGGGACAGGTACTTAGACGAC
 TGAACCTGGCCTCTGTTCTAATGGTTGAGCCTCAGAGTACATATTGGGTGCGGTTTG
 GTTGCTTGAGAGTTGGTTGTCACATGTGATGTTCTGCCATTCCAGGACGA
 AAAGGAGGTTGTGCTAAAGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTAC
 [G, A]
 GGGCAATGGCATCGTGGAGCGCAGGTAGTCAGGACGAGGATTGTGCTGCTTCTAGA
 TTTGGCCCTGGACTTTGTCACAGCCCCAGACAACCTTCTGAAGC
 TAGTACAGCCATGTGCTGCACAGTGACGCTTGGTCAATGTCGATATATGATGTTGGAC
 CCATAAGATTATAATGGAGCTGAAAATTCCGTGCGCTAGTGTATGTTGACAA
 CACATTACCTTTCTACGTTAGGTACACAAATATTGCCCTACAGGATTCACTAGAGTC

 11135 AGGAAACTAAGTGAAGTAGAGAGTAGGAAAAGCTGGAAAGGGTTGAAGCACATGGAACTAGT
 GAAAGGACAAGGAGAACATGTGTTGCCTGGAGGGACAGGTACTTAGACGACTGAAC
 GCCTGTTCTAATGGTTGAGCCTCAGAGTACATATTGGGTGCGGTTGGTTGCT
 TTGTAGAGTTGGTTGTCACATGTGATGTTCTGCCATTCCAGGACGAAAAGGAG
 GTTGTGCTAAAGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTACGGGGCA
 [A, G]
 TGGCATCGTGGAGCGCAGGTAGTCAGGACGAGGATTGTGCTGCTTCTAGATTGGCC
 CTGGACTTGTCTGAAAAACCTTCACAGCCCCAGACAACCTTCTGAAGCTAGTACA
 GCCATGTGCTGCACAGTGACGCTTGGTCAATGTCGATATATGATGTTGGACCCATAAG
 ATTATAATGGAGCTGAAAATTCTGTGCGCTAGTGTATGTTGAGTGGCACAACACATTA
 CCTTTCTACGTTAGGTACACAAATATTGCCCTACAGGATTCACTAGAGTCACATGCT

 11143 AAGTGAAGAGAGTAGGAAAAGCTGGAAAGGGTTGAAGCACATGGAACTAGTGAAGGAC
 AAGGAGAACATGTGTTGCCTGGAGGGACAGGTACTTAGACGACTGAACCTGGCTCTGT
 GTTCTAATGGTTGAGCCTCAGAGTACATATTGGGTGCGGTTGGTTGCTTAGAG
 TTGGTTGTCACATGTGATGTTCTGCCATTCCAGGACGAAAAGGAGGTTGTGCT
 AAAGGAAAGGAGGATCGATGCACATGTATGCCAAGAACTTCTACGGGGCAATGGCATC
 [G, A]
 TGGAGCGCAGGTAGTCAGGACGAGGATTGTGCTGCTTCTAGATTGGCCCTGGACCT
 TGTCTGAAAAACCTTCACAGCCCCAGACAACCTTCTGAAGCTAGTACAGCCATGTG
 CTGCACAGTGACGCTTGGTCAATGTCGATATATGATGTTGGACCCATAAGATTATAAT
 GGAGCTGAAAATTCTGTGCGCTAGTGTATGTTGAGTGGCACAACACATTACCTTCT
 ACGTTAGGTACACAAATATTGCCCTACAGGATTCACTAGAGTCACATGCTGTGAGGG

 12486 TTACAGCGTGAGCCACCACGCCGGCCAGGTTTATTCTTAACCTTGATGCAGAAA
 TGTTAGTCTTACTGGTAAAATAGAACATAGTATTATATTACTTTAGTGTCTTATT
 GAAAATATCGGAGGTTGGATAAACAGAGAGATAGGGTTGAGGAGAGTTGTAGCAGCA
 GTGTAATTCTGTGTCAGATTCTGCCAGGAGTGAATGCAAGGGCATTAAATTAGTATCT
 CCCCTCATGGATTCTGTGGTTCTTCTCGGTTGCTTAATGTTAGGTGCCCTGGG
 [G, C]
 CTGGGATTGCTCTAGCCTGTAAGTATAATGGAAAAGATGAGGTCTGCCCTGACTTTATATG
 GCGATGGTGTGCTAACCCAGGTAAATTATGTCCTTAACCTCCAAAACAGTCTTATT
 CAAAGTCTTAATATTACAGTGAATTCTAAAGAAGTAGCATATTGCTTATTAGGTGA
 AATAGCAAGTCTATGGCTAGCTCAAATTGGTTGACTTATGCCAGATTAGAGATTGAC
 CTCTTAGGTTGTTACAAGAGACTACGGGGCACATTCTGTGAAGGAGCTCACCTT

 12558 CTGGTTAAAATAGAACATAGTATTATATTACTTTAGTGTCTTATTGAAAATATCGGA
 GGTGGGATAAACAGAGAGATAGGGTTGAGGAGAGTTGTAGCAGCAGTGTAAATTCTG
 TGTCAGATTCTGCCAGGAGTGAATGCAAGGGCATTAAATTAGTATCTCCCTCATGGAT
 TTCTGTGGTCTTCTCGGTTGCTTAATGTTAGGTGCCCTGGCGCTGGGATTGCT
 CTAGCCTGTAAGTATAATGGAAAAGATGAGGTCTGCCCTGACTTTATATGGCAGTGGCT
 [G, A]
 CTAACCAGGTAAATTATGTCCTTAACCTCCAAAACAGTCTTATTGAAAGTCTTAA
 TATTACAGTGAATTCTAAAGAAGTAGCATATTGCTTATTAGGTGAATAGCAAGTCC
 TATGGCTAGCTCAAATTGGTTGACTTATGCCAGATTAGAGATTGACCTTACGGTTG
 TTTCACAAGAGACTACGGGGCACATTCTGTGAAGGAGCTCACCTTGCTACATCA
 GTGCTTGGCAAAGGCCCTGTGGTAAAGGACCTCCCCACACCTATTGCAAAACAATACAG

 13376 TCCTTACTGCTAGCATGGCCCCAAAGAAACAAGTAGTAGTTGGTTGTCACCTCCTTAG
 TTGCAAGAGTATGATGCCCTGCTACTTCTCCACCACCCACCCGCTTCCCTCAC

FIGURE 3J

CCAAAGCTCGGTTTAGAAGAGGAGGCTTCTGTGCTTATGAAAGCTTCTGTGCCAGG
 CAGAGCAGCAGCTTAGAGATGATGAAGCCTGGAGAAAGAAGCCAAATGAAACCCCTT
 TCGTAACTACTTCCAGGGCCAGATATTGAGCTTACAACATGGCAGTTGTGAAATT
 [T, C, A]
 CCTTGTATTTCATCTGTGAGAATAATCGCTATGGAATGGAACGTCTGTT

13378
 CTTACTGCTAGCATGGCCCCAAAGAAACAAGTAGTAGTTGGTTGTACCTTCCTTAGTT
 GCAAGAGTATGATGCTGCTACTTCTCCACCAACCCACCCGCTTCCCTCACCAACCC
 AAAGCTCGGTTTAGAAGAGGAGGCTTCTGTGCTTATGAAAGCTTCTGTGCCAGGCA
 GAGCAGCAGCTTAGAGATGATGAAGCCTGGAGAAAGAAGCCAAATGAAACCCCTTTC
 GTAACTACTTCCAGGGCCAGATATTGAGCTTACAACATGGCAGTTGTGAAATTAC
 [C, T]
 TTGTATTTCATCTGTGAGAATAATCGCTATGGAATGGAA

16233
 GTGGGACAAGTGCACACTCTTTAGAACCTTACATCTAAATCAAAGCAGCAAGCAAAA
 ACTGGCCCTGTTGCGTAATGCCAGGGAAAGCCATGTGACTCACAGTGTACGGTTTCTA
 GAAAAGACAGAAGCAGTTTACAGAATGTTAGGCTGCGTTCTGGTATTTGAAAGTATA
 ACAACAACCTGCCACGCCATAGTGACATAAGCATTGGTATGCCCTTGTGTTCAGAAA
 CACACTCTGTATTCACCTCATTGGACAATCCAACCCATATCATGTTCATCACGCC
 [G, C]
 TCCTTGCTCTACTGGAACTGCTTACTGATGATTACTACTTTCCCTCCCATAGTTA
 CCGTACACGAGAAATTAGGAAGTAAGAAGTAAGAGTGACCCATTATGCTTCTCAA
 GGACAGGATGGTGAACAGCAATCTGCCAGTGAGAAGACTAAAGGTACAGTCACTTGT
 TCATGGTGGTTGAAGGTTGGTTAAAAGTTGCCACCCCTGGGTGGCCACAGAGTTGT
 GTGGGTTCTCCAAGCCCAGAAAGTGATGTCCTGGGACATAATAGTCCATAGTCCAA

16354
 AAAAGACAGAACGAGCTTATTACAGAATGTTAGGCTGCCITCTGGTATTTGAAAGTATAA
 CAACAACCTGCCACGCCATAGTGACATAAGCATTGGTATGCCCTTGTGTTCAGAAC
 ACACCTCTGTATTCACCTCATTGGACAATCCAACCCCATATCATGTTCATCACGCC
 TCCTTGCTCTACTGGAACTGCTTACTGATGATTACTACTTTCCCTCCCATAGTTA
 CCGTACACGAGAAATTAGGAAGTAAGAAGTAAGAGTGACCCATTATGCTTCTCAA
 [G, A]
 GACAGGATGGTGAACAGCAATCTGCCAGTGAGAAGAACTAAAGGTACAGTCACTTGT
 CATGGTGGTTGAAGGTTGGCTTAAAAGTTGCCACCCCTGGGTGGCCACAGAGTTGT
 TGGGTTCTCCAAGCCCAGAAAGTGATGTCCTGGGACATAATAGTCCATAGTCCAA
 GTCCCTGGGGTGGGGCTTCTTAGTTCCCTATTCAAAATTGTATTACTCTTCA
 GATTCAGATTTGGTGGACTGTGAACCACCATCACAGTGGCAAAGCCCCACAGTAGTA

16377
 GAATGTTAGGCTGCCCTGGTATTTGAAAGTATAACAACAACTCTGCCACGCCATAG
 TGACATAAGCATTGGTATGCCCTTGTGTTCAAGAAACACACTCTGTATTCACCTCATT
 GGGACAATCCAACCCATATCATGTTCATCACGCCGCTTGTCTACTGGAACTGCTC
 TTACTGATGATTACTACTTTCCCTCCCATAGTTACCGTACAGGAGAAGAAATTGAG
 AAGTAAGAAGTAAGAGTGACCCATTATGCTCTCAAGGACAGGATGGTAACAGCAATC
 [T, G]
 TGCCAGTGTGAAAGAACTAAAGGTACAGTCATTGTTATGGGGTTGAAGGTTGGCTT
 TAAAAGTTGCCACCCCTGGGTGGCCACAGAGTTGTGTTGGCTTCTCCAAGGCCAGAAAG
 TGATGTCCTGGGACATAAAATTGATTCATAGTCCAAAGTCCCTGGGTGGGGCTTTC
 CTTTAGTTCTATTCAAAATTGTATTACTCTTCAGATTTCAGATTGGTGGACTGT
 GAACCACCATCACAGTGGCAAAGCCCCACAGTAGTATGGTCTTTCTAAAGTAT

FIGURE 3K